Nucleic Acids Research, 1998, Vol. 26, No. 15 **3567–3571**

Solution studies of the dimerization initiation site of HIV-1 genomic RNA

Frédéric Dardel*, Roland Marquet[1], Chantal Ehresmann[1], Bernard Ehresmann[1] and Sylvain Blanquet

Laboratoire de Biochimie, UMR 7654 du CNRS, Ecole Polytechnique, 91128 Palaiseau Cedex, France and [1]IBMC, UPR 9002 du CNRS, 15 rue René Descartes, 67084 Strasbourg Cedex, France

Received March 12, 1998; Revised and Accepted June 1, 1998

ABSTRACT

Dimerization of HIV-1 genomic RNA is an essential step of the viral cycle, initiated at a conserved stem–loop structure which forms a 'kissing complex' involving loop–loop interactions (dimerization initiation site, DIS). A 19mer RNA oligonucleotide (DIS-19) has been synthesized which forms a stable symmetrical dimer in solution at millimolar concentrations. Dimerization does not depend on addition of Mg^{2+}. RNA ligation experiments unambiguously indicate that the formed dimer is a stable kissing complex under the NMR experimental conditions. ^{1}H NMR resonance assignments were obtained for DIS-19 at 290 K and pH 6.5. Analysis of the pattern of NOE connectivities reveals that the helix formed by loop–loop base pairing is stacked onto the two terminal stems. Non-canonical base pairs between two essential and conserved adenines are found at the junctions between the two intramolecular and the single intramolecular helices.

INTRODUCTION

HIV-1, like all retroviruses, has a diploid genome composed of two identical RNA molecules. It has been shown that these two RNA strands are bound through a small region located close to their 5′-terminus. This dimerization is essential for several steps of the retroviral replicative cycle, including encapsidation and reverse transcription (reviewed in 1). RNA fragments corresponding to the 5′-end of HIV-1 genomic RNA retain the ability to dimerize *in vitro* (2). Further studies by chemical interference indicated that initiation of dimerization involves a conserved stem–loop structure around position 275, called the dimerization initiation site (DIS) (3). This stem–loop structure features a palindromic hexanucleotide sequence within the loop, GUGCAC or GCGCGC, which has been shown to mediate dimerization via a 'kissing complex' involving loop–loop interactions (3–7). In the case of the HIV-1_{Mal} isolate, it has been additionally shown that the stem sequence can be exchanged without significant modification of dimer stability, thereby suggesting that the dimer remains in the kissing complex state (8). Mutations of the DIS affect at least two steps of the retroviral cycle: RNA packaging (9–11) and synthesis of proviral DNA (10). The DIS is thus a potentially interesting target for antiviral agents. In this context, it is of great interest to obtain structural information on initial dimer formation. Based on site-directed mutagenesis and chemical modification studies, a model has recently been proposed. It includes a number of non-canonical base interactions and accounts for the high stability of the RNA dimer (12). The present work presents a ^{1}H NMR analysis of dimerization and the solution conformation of a short RNA oligonucleotide corresponding to HIV-1 DIS by ^{1}H NMR. The reported results provide direct experimental support for the model.

MATERIALS AND METHODS

RNA transcription and purification

Template oligonucleotides (5′-GGCTGTGTGCACCTCAGCCTA-TAGTGAGTCGTATTA and 5′-TAATACGACTCACTATAG) were purchased from Eurogentec (Belgium). *In vitro* transcriptions using purified T7 RNA polymerase were performed according to the method of Milligan and Uhlenbeck (13). The preparative scale experiment was conducted in a 250 ml reaction containing 2 mM each NTP and 200 nM annealed template oligonucleotides. After phenol extraction and ethanol precipitation, the RNA transcripts were dissolved in 20 mM sodium phosphate, pH 7.0, and purified by two steps of ion exchange chromatography on a Mono-Q column (Pharmacia), using a 200–800 mM linear NaCl gradient in the above buffer. Under these conditions, the desired transcript could be separated from most of the n + 1 by-product. The final yield was 24 mg purified RNA.

UV melting curves were recorded using a Uvikon 922 spectrophotometer (Kontron) equipped with a programmable temperature control unit. The RNA sample was dissolved in 20 mM sodium phosphate, pH 6.5, at concentrations ranging from 0.5 to 5 μM. After denaturation at 95°C, the sample was transferred to a spectrophotometer and the temperature was slowly varied from 80 to 20°C and back to 80°C (rate 0.5°C/min). Melting curves were analysed with the manufacturer's software.

RNA ligation

Calf intestine alkaline phosphatase (molecular biology grade), polynucleotide kinase and T4 RNA ligase were from Boehringer Mannheim. All reactions were carried out using the buffer

*To whom correspondence should be addressed. Tel: +33 1 69 33 41 81; Fax: +33 1 69 33 30 13; Email: fred@botrytis.polytechnique.fr

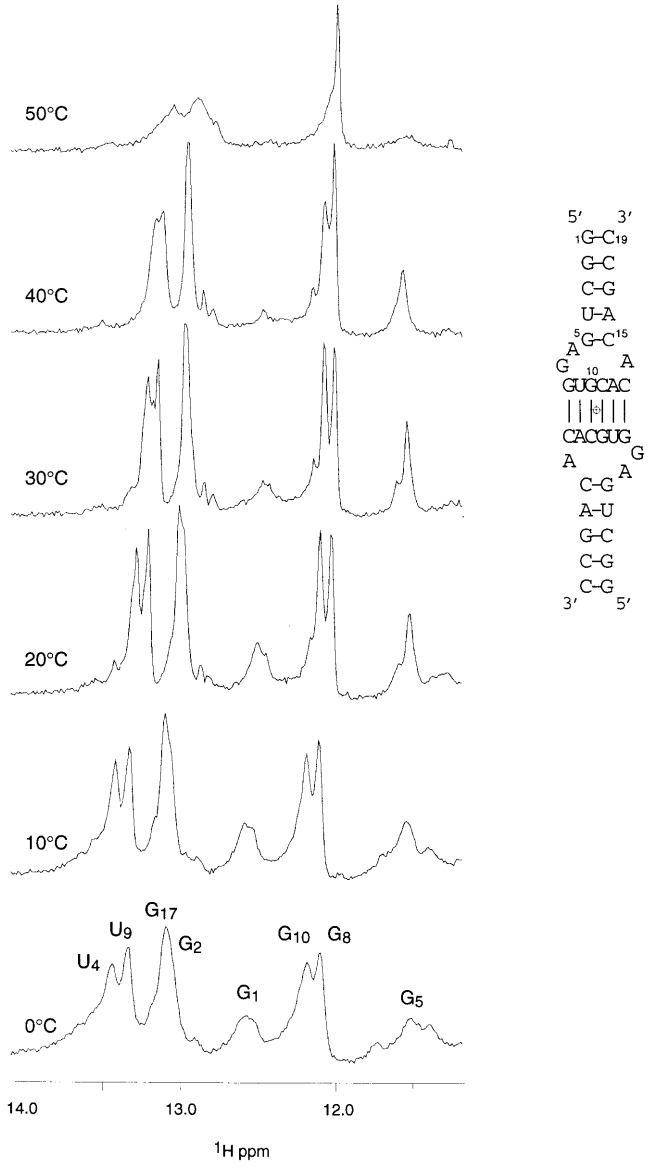

Figure 1. Denaturation of 1 mM DIS-19 dimer (20 mM sodium phosphate, pH 6.5) followed by 1D NMR. The region of the spectrum corresponding to the base paired imino protons is shown. Eight resonances are observed, corresponding to 16 bp related by a 2-fold symmetry (right).

conditions recommended by the supplier. The RNA transcript (30 nmol/100 µl) was treated with 2 U phosphatase for 2 h at 37°C. After phenol extraction and ethanol precipitation, the dephosphorylated RNA was resuspended in 100 µl and 5′-end-labelled using polynucleotide kinase (10 U) and [γ-^{32}P]ATP (1.85 MBq; Dupont-NEN). Phosphorylation of the 5′-end of the transcript was then completed by further incubation (1 h), following addition of 0.4 mM unlabelled ATP. The RNA was finally phenol extracted, ethanol precipitated and resuspended in H_2O.

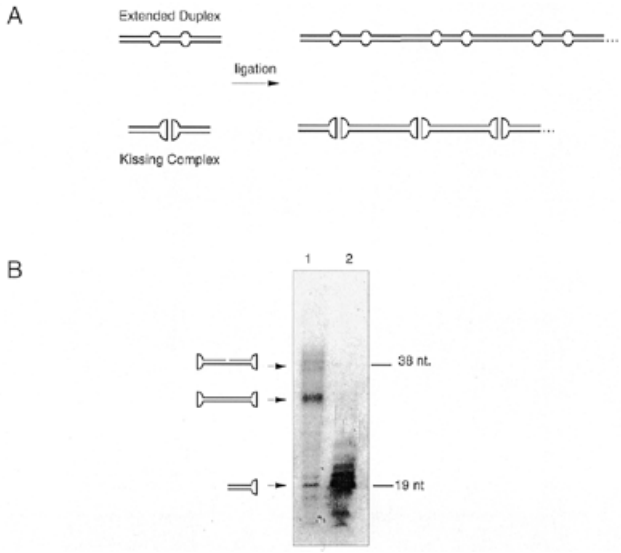

Figure 2. Ligation of DIS-19 dimer. (**A**) Upon ligation, an extended duplex (top) and a kissing complex (bottom) will give different end products which can be distinguished by denaturing gel electrophoresis. In the former case, high order concatemers will be generated, whereas in the latter, only cyclic dimers associated by non-covalent interactions will be produced. (**B**) DIS-19 ligation. Lane 1, ligation end products; lane 2, unligated DIS-19. The position of unlabelled size markers is indicated on the right.

Ligation experiments were conducted in 10 µl reactions, containing 4 µg ^{32}P-labelled RNA (~60 µM) and 40 U T4 RNA ligase. After overnight incubation at 37°C, the reaction was essentially complete and no further change in the pattern of products was observed, even upon addition of further ligase and ATP. The products were separated by electrophoresis on 12% polyacrylamide–8 M urea gels and revealed by autoradiography.

NMR experiments

The NMR samples were prepared in 20 mM potassium phosphate (pH 6.5), 0.02% NaN_3 (w/v), in either $H_2O/^2H_2O$ (9:1 v/v) or 100% ^2H_2O. 2D NMR spectra were recorded at 290 K on Bruker AMX600 and DRX600 spectrometers, using a 2 mM RNA sample. 2D homonuclear experiments were recorded and processed as previously described (14). DQF-COSY, TOCSY (mixing time 55 ms) and NOESY experiments (mixing times 75, 120 and 200 ms) were recorded over spectral widths of 5556 Hz for experiments recorded in ^2H_2O and 11 111 Hz for those recorded in H_2O. In order to resolve some ambiguities in the assignment process caused by spectral overlap, additional NOESY spectra (mixing time 120 ms) were also collected at 308 and 278 K.

RESULTS AND DISCUSSION

DIS-19 folds into a single stable dimeric form

As a first step towards a structural analysis of the HIV-1 dimerization process, we investigated the ability of 'minimal' RNA fragments containing the DIS to form stable dimers *in vitro*. Of several oligoribonucleotides analysed for their stability and for dimerization (not shown), we selected the following sequence: 5′-GGCUGAGGUGCACACAGCC-3′, hereafter called DIS-19. This RNA fragment corresponds to nt 268–283 of HIV-1_{Mal},

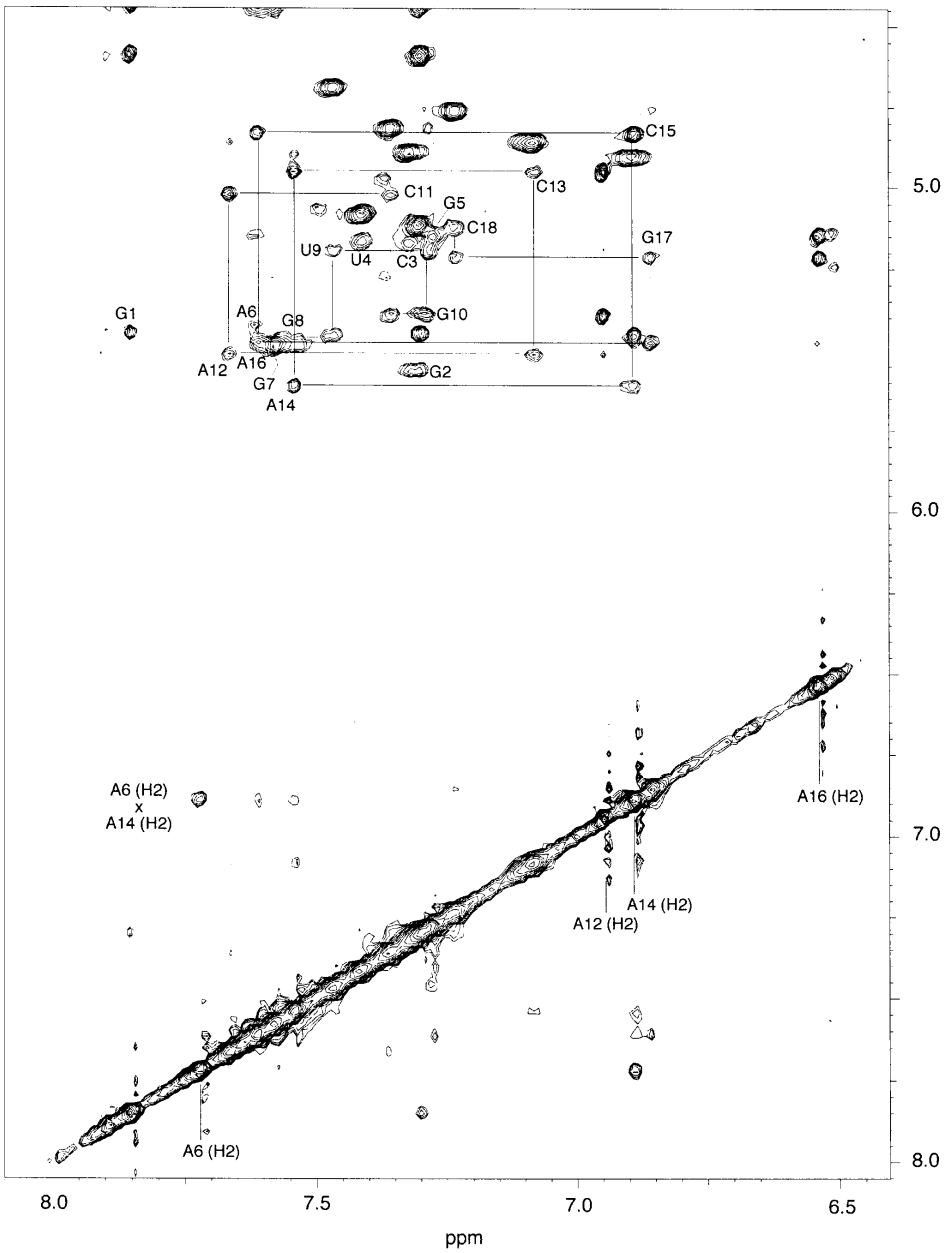

Figure 3. Sequential resonance assignment of DIS-19. Shown is a portion of a 2D NOESY spectrum recorded in ^2H_2O (pH 6.5, 290 K, mixing time 120 ms). Internal H1′–H6/H8 crosspeaks are labelled. A sequential 'walk' spanning G8 to C18 is shown. The H2–H2 crosspeak between A6 and A14 is indicated.

closed by an additional G-C base pair (underlined in the sequence), and has been synthesized in large quantities by *in vitro* transcription, as described in Materials and Methods. The DIS-19 sequence can be folded into a stem–loop structure. The loop exposes a palindromic GUGCAC sequence which could promote dimerization via a loop–loop 'kissing' interaction (Fig. 1). A preliminary analysis of the conformation of DIS-19 in solution was performed by 1D NMR spectroscopy. In the imino proton region, eight resonances corresponding to base pairs can be observed (Fig. 1), indicating that, at millimolar concentrations, DIS-19 folds as a symmetrical dimer even at low ionic strength. When melting of DIS-19 was followed by 1D NMR, all of the above imino resonances remained detectable up to 50°C and disappeared simultaneously at 60°C. This behaviour is in good agreement with a melting temperature of 54 ± 3°C determined by UV spectroscopy. It also indicated that melting of DIS-19 was cooperative and that, in particular, opening occurred at the same time for the base pairs between the two central palindromic sequences and for those in the two terminal stems. Unfolding is fully reversible, as evidenced by restoration of the initial NMR spectrum after cooling the sample below the melting temperature. The influence of ionic strength and divalent cations on the structure of the DIS-19 dimer was also investigated. 1D and 2D NMR spectra were mostly unaffected by addition of either 100 mM NaCl or 10 mM $MgCl_2$, except for an increase in the linewidth. DIS-19 thus folds into a unique stable dimeric form.

Table 1. ^{1}H chemical shift assignments of DIS-19 at pH 6.5 and 290 K

	Imino	Amino	H6/H8	H5/H2	H1′	H2′	Other ribose
G1	12.50		7.81		5.41	4.55	4.39, 4.15, 4.06, 3.89
G2	13.03		7.26		5.52	4.18	4.14
C3		6.48, 8.18	7.28	4.85	5.13	4.05	
U4	13.32		7.37	5.04	5.11	4.08	
G5	11.50		7.24		5.12	4.17	
A6			7.58	7.68	5.39	4.40	
G7	10.18		7.53		5.45	4.37	
G8	12.11	6.14	7.49		5.41	4.24	
U9	13.25		7.43	4.65	5.16	4.22	4.15, 4.06
G10	12.04	5.88	7.25		5.36	4.01	4.17, 4.08
C11		6.34, 7.92	7.31	4.78	4.98	4.01	4.12
A12			7.62	6.91	5.47	4.15	4.25, 4.06, 4.01
C13		6.54, 7.63	7.04	4.82	4.91	4.06	4.00
A14			7.50	6.85	5.57	4.04	4.09
C15		6.59, 7.88	6.86	4.86	4.79	3.84	4.06, 3.95
A16			7.57	6.50	5.44	4.20	4.03
G17	13.00		6.83		5.17	3.99	
C18		6.47, 8.14	7.19	4.72	5.08	3.83	3.95
C19		6.60, 8.01	7.26	5.07	5.34	3.59	3.75

Chemical shifts are indicated in parts per million (p.p.m.), relative to an internal trimethylsylyl propionate reference.

DIS-19 dimer is a kissing-loop complex and not an extended duplex

In principle, a quasi-palindromic RNA oligonucleotide, such as DIS-19, can dimerize either as a 'kissing complex', schematized in Figure 1, or as an extended duplex. Stem exchange experiments strongly suggested that, at least in the case of HIV-1_{Mal}, the dimer remained in the kissing complex form (8). Both the extended duplex and the kissing complex contain 16 bp with a central 2-fold axis of symmetry and NMR experiments cannot easily distinguish between the two forms. In order to establish the dimerization mode of DIS-19, we therefore developed an original strategy, using RNA ligation to probe the fold of this RNA oligonucleotide. DIS-19 was 5′-end-labelled and then incubated at a high concentration (60 μM), as close as possible to that used in the NMR studies, in the presence of T4 RNA ligase and ATP. The joining of RNA extremities by this enzyme should yield very different end products for the 'kissing complex' and the extended duplex. In the latter case, ligation should result in formation of oligomers and should be driven towards formation of high order covalent multimers of the original RNA oligonucleotide (Fig. 2A, top). In the case of the kissing complex, the final product is expected to be a cyclic dimer, where the bottom of the stems of two individual hairpins have been ligated together (Fig. 2A, bottom), and no higher order covalent oligomers should be observed. Ligations of DIS-19 were thus allowed to go to completion and the resulting products were analysed by electrophoresis on a denaturing gel. A typical result is shown in Figure 2. A small amount of linear dimer is observed, but no higher order oligomers can be detected, even using much longer exposure times (not shown). Instead, the major product is reproducibly a species that migrates slightly faster than the linear dimer and should thus correspond to the cyclic dimer. This conclusion is supported by the observation that the radioactive phosphate groups carried by this species have become fully resistant to alkaline phosphatase treatment (not shown), as expected for a cyclic molecule. This experiment strongly suggests that, as previously proposed, DIS-19 remains in the 'kissing complex' state. The small amount of linear dimer observed could result either from incomplete 5′-phosphorylation of the sample or from a slight degradation of DIS-19 termini during the dephosphorylation and phosphorylation steps, both of which would prevent closure of the dimer.

^{1}H NMR resonance assignments and solution conformation of DIS-19

Most assignments were obtained at 290 K, a temperature at which the imino resonances were resolved and sharp. All could be assigned by a combination of 1D NOE difference and 2D NOESY experiments. The two uracil imino protons were readily identifiable from the strong NOE to the H2 of the facing adenine. The ribose H1′–H2′ ^{3}J couplings of all nucleotides, except A6 and G7, were very small, indicating that they have a classical 3′-*endo* conformation, typical of A-type RNA helices. Indeed, continuous H1′–H6/H8 sequential walks could be observed for G1 to A6, on the one hand, and from G8 to C19, on the other hand (Fig. 3). This indicated that the central kissing loop helix is stacked onto the two stem helices, at least on the 3′-side (from G8 to C19). This rendered the assignment procedure almost straightforward using the classical strategy for oligonucleotide duplexes (15), since the H1′–H6/H8 region is well resolved (Fig. 3). A

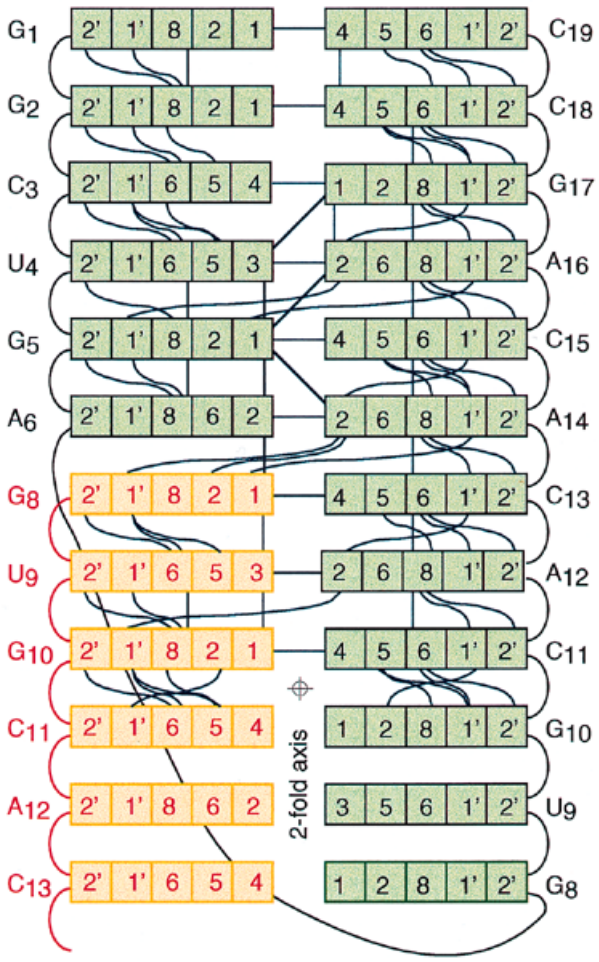

Figure 4. Summary of the observed internucleotide NOE connectivities. Each numbered square represents the corresponding proton of the given nucleotide. Grey letters and boxes correspond to a portion of the second RNA strand in the dimer. Only half of the connectivities are indicated, since the others are symmetry related.

summary of the observed internucleotide connectivities is shown in Figure 4 and the deduced chemical shifts are listed in Table 1.

Two other kissing loop structures have been reported, that of the HIV TAR:TAR* complex (16, 17) and of RNA I:RNA II of the ColEI plasmid (18). In both cases, stacking of the central loop and of the two helices is also observed, with the kink in the loop also being located on its 5′-side, as for the DIS dimer. The major difference between DIS and both the TAR and ColEI complexes lies in the size of the loop, which is 9 nt long in DIS, compared with 6 nt for TAR and 7 nt for ColEI. This longer loop in DIS could result in a less constrained loop–loop interaction and might thus explain why the symmetrical kissing complex appears favoured over the extended duplex.

Two conserved purines, A6 and A14, which flank the self-complementary hexanucleotide of the loop, have been reported to be important for DIS dimerization (12,19). Interestingly, both adenines are stacked inside the helix, as shown from the numerous NOEs observed with the adjacent bases (G5, C15 and C13, Fig. 4), and thus face each other, forming a non-canonical A:A base pair, as evidenced by a non-ambiguous NOE H2–H2 contact between the two bases (Fig. 3). These two observations, (i) stacking of the three helical stems and (ii) formation of an A:A base pair, are in good agreement with the model previously proposed based on chemical protection data (12). The position of G7, however, could not be established, since little or no NOEs are seen from the protons of this base, in part because its exchangeable protons exhibit broad resonances. The accurate determination of the 3D structure of DIS-19 will require isotopic enrichment of the RNA with ^{13}C and ^{15}N, which is currently under way.

ACKNOWLEDGEMENTS

The authors acknowledge Ms Manuela Cuffiani for her contribution to the initial stage of this work and Prof. Jean-Yves Lallemand for access to the NMR facility of the Departement de Chimie of the Ecole Polytechnique. This work was supported by the Agence Nationale de Recherche sur le Sida.

REFERENCES

1 Paillart,J.C., Marquet,R., Skripkin,E., Ehresmann,C. and Ehresmann,B. (1996) *Biochimie*, **78**, 639–653.
2 Marquet,R., Paillart,J.C., Skripkin,E., Ehresmann,C. and Ehresmann,B. (1994) *Nucleic Acids Res.*, **22**, 145–151.
3 Skripkin,E., Paillart,J.C., Marquet,R., Ehresmann,B. and Ehresmann,C. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4945–4949.
4 Paillart,J.C., Marquet,R., Skripkin,E., Ehresmann,B. and Ehresmann,C. (1994) *J. Biol. Chem.*, **269**, 27486–27493.
5 Laughrea,M. and Jetté,L. (1994) *Biochemistry*, **33**, 13464–13474.
6 Muriaux,D., Girard,P., Bonnet-Mathonnière,B. and Paoletti,J. (1995) *J. Biol. Chem.*, **270**, 8209–8216.
7 Clever,J.L., Wong,M.L. and Parslow,T.G. (1996) *J. Virol.*, **70**, 5902–5908.
8 Paillart,J.C., Skripkin,E., Ehresmann,B., Ehresmann,C. and Marquet,R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 5572–5577.
9 McBride,M.S. and Panganiban,A.T. (1996) *J. Virol.*, **70**, 2963–2973.
10 Paillart,J.C., Berthoux,L., Ottmann,M., Darlix,J.L., Marquet,R., Ehresmann,B. and Ehresmann,C. (1996) *J. Virol.*, **70**, 8348–8354.
11 Berkhout,B. and van Wamel,J.L. (1996) *J. Virol.*, **70**, 6723–6732.
12 Paillart,J.C., Westhof,E., Ehresmann,C., Ehresmann,B. and Marquet,R. (1997) *J. Mol. Biol.*, **270**, 36–49.
13 Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.
14 Wallis,N.G., Dardel,F. and Blanquet,S. (1995) *Biochemistry*, **34**, 7668–7677.
15 Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York, NY.
16 Chang,K.Y. and Tinoco,I.,Jr (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 8705–8709.
17 Chang,K.Y. and Tinoco,I.,Jr (1997) *J. Mol. Biol.*, **269**, 52–66.
18 Marino,J.P., Gregorian,R.S.,Jr, Csankovski,G. and Crothers,D.M. (1995) *Science*, **268**, 1448–1454.
19 Skripkin,E., Paillart,J.C., Marquet,R., Blumenfeld,M., Ehresmann,B. and Ehresmann,C. (1996) *J. Biol. Chem.*, **271**, 28812–28817.